An Indexing Standard for Sharing Health Education Multimedia Resources: The Health Education Assets Library (HEAL) Metadata Schema

Sharon E. Dennis, M.S.
Spencer S. Eccles Health Sciences Library, University of Utah

Chris S. Candler, M.D.
The University of Oklahoma College of Medicine

Sebastian Uijtdehaage, Ph.D.
David Geffen School of Medicine at UCLA

Sandra A. McIntyre, M.Ed.
David Geffen School of Medicine at UCLA

Shona Dippie, M.L.S.
Spencer S. Eccles Health Sciences Library, University of Utah

Corresponding Author
Sharon E. Dennis, M.S.
Librarian for Multimedia Development
Spencer S. Eccles Health Sciences Library
University of Utah
10 North 1900 East
Salt Lake City, UT 84108
Phone: (801) 585-3928
Fax: (801) 581-3632
sdennis@lib.med.utah.edu

An Indexing Standard for Sharing Health Education Multimedia Resources: The Health Education Assets Library (HEAL) Metadata Schema

Abstract

Health sciences educators are increasingly incorporating multimedia (including images, animations, and videos) into educational materials such as Powerpoint lectures, Web sites, and interactive quizzes and cases. Educators continue to "reinvent the wheel" and develop costly duplicates of multimedia resources, despite new opportunities offered by the Internet to share resources. The Health Education Assets Library (HEAL) is designed to provide health sciences educators with freely available, high-quality multimedia materials to augment health sciences education. We describe an XML schema that we created to index the health sciences multimedia resources in HEAL. The metadata schema provides a common mechanism by which remotely located distributed systems may share metadata records, allowing the end user to search many collections through one interface.

Introduction

Rapid advances in technology have created new opportunities for health sciences educators to integrate multimedia resources into the curriculum. A high quality image with annotations, an animation depicting a difficult concept, or a video of a lecture that can be reviewed "on demand" can add to the educational experience for both faculty and students. Web-based course management systems now make it possible for faculty to offer students a wide variety of course materials in a customized environment. Faculty can offer students interactive experiences, through online cases and quizzes, that are not possible with traditional educational methods. The use of educational technology by health sciences educators has increased steadily for years; as the Internet has become ubiquitous within the last few years, use has increased dramatically. [1 2 3 4 5] Despite the increased use of technology in health sciences education, many faculty continue to "re-invent the wheel" by creating "yet another" brain atlas or video archive of neurologic signs and symptoms. This occurs either because the faculty member is unaware of similar resources or does not have access to re-purposeable materials that they may adapt to their particular curricular needs and teaching style. The need for establishment of consortia and standards that would facilitate collaborative development and sharing of educational resources is apparent.[6 7 8] Educators would obviously benefit from an international multimedia repository that would offer "one-stop shopping" for freely available, high quality multimedia resources in a variety of subject areas.

Methods

In 2000 the authors received an award from the National Science Digital Library Initiative (NSDL)[9] of the National Science Foundation (NSF) to plan, design, and develop a web-based application to facilitate sharing of healthcare education materials.[10] The Health Education Assets Library (HEAL)[11] was subsequently established as a core component of the NSDL whose purpose is to provide educators with free, high-quality multimedia materials to augment health sciences education. The goals of the HEAL project are fourfold: i) to improve access to high-quality multimedia resources, ii) to promote sharing of educational resources between faculty and institutions, iii) to foster interoperability through the creation of standards, and iv) to establish a sustainable infrastructure. Through collaboration with numerous faculty, medical schools, the Association of American Medical Colleges, and the National Library of Medicine, the HEAL project desires to meet the educational resource needs of healthcare educators across the educational continuum.

The HEAL application is available to the public at our Web site (http://www.healcentral.org). An initial prototype collection of 2500 items was made available beginning June 2003. In 2002, the authors received a second NSF grant to substantially increase the breadth and depth of the HEAL collection. We anticipate offering upwards of 50,000 items by the fall of 2004.

The HEAL application allows users to browse the collection using a hierarchical tree of controlled vocabulary (Figure 1) or to conduct a simple keyword search . In addition, users may conduct an advanced search that includes options for narrowing a search by author, subject heading, or health science specific characteristics (specimen type, radiograph type, etc.) (Figure 2). The search results page (Figure 3) includes a

thumbnail depiction of each item in addition to a description, the source collection, the format, and the file size; users can add an item to a download folder for later batch downloading**.**

While the HEAL digital library accepts contributions from individuals on our Web site, we recognized that the HEAL collection would not cover a wide variety of subject areas without a proactive collection development policy. To that end, we formed partnerships with other institutions and individuals that had already developed high quality collections of materials. To date over 100 partners have expressed an interest in contributing a collection to HEAL. Since uploading multimedia materials from these external collections onto the central HEAL server would be expensive and impractical, the HEAL system would need to identify a mechanism to connect and search remote collections from a single system.

Using the Open Archives Initiative (OAI)[12] specification, a standard designed to facilitate the interchange of catalog information between web-based collections, we identified and tested a federation model that could support the exchange of catalog information between the central HEAL server and the remote partner collection servers. Such an approach allows partner collections to include their catalog information into the central HEAL catalog independent of their desire to contribute some or all of their multimedia resources (Figure 4).

A consistent standard for describing each multimedia resource in the HEAL collection is necessary in order for users to be able to retrieve relevant items from the database as well as for sharing information between the distributed remote collections. The descriptive information about each item is referred to as "metadata" (often defined as

information that describes information). A metadata standard consists of a precise set of descriptive elements; for instance, "title" to describe the title of the resource, "author" to describe the author, etc. A common standard is necessary to ensure that there is no confusion regarding the meaning of each element; for instance, an element called "title" could be interpreted in multiple ways (i.e., a faculty “title” vs. the “title” of the resource). Standards improve the quality of the metadata, provide a mechanism for keeping the metadata consistent across systems, help to ensure that the resources can be accessed in the future, and facilitate the exchange of metadata between disparate systems.[13]

Metadata Significance

The idea of describing resources with metadata, while not a new concept, takes on added importance in today's distributed digital environment. As Weibel and Lagoze note, "The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself."[14]

A metadata record consists of a set of attributes, or elements, necessary to describe the resource in question. For example, a metadata system common in libraries -- the library catalog -- contains a set of metadata records with elements that describe a book or other library item: author, title, date of creation or publication, subject coverage, and the call number specifying location of the item on the shelf.

The linkage between a metadata record and the resource it describes may take one of two forms: elements may be contained in a record separate from the item, as in the case of the library's catalog record; or the metadata may be embedded in the resource itself. Many metadata standards in use today, including the Dublin Core standard[15], do not prescribe either type of linkage, leaving the decision to each particular implementation. The HEAL application separates the multimedia content from the metadata description. If the multimedia item is located on a remote server, the metadata for the item is periodically harvested into the central HEAL server through the OAI protocol described previously. This allows the owner of the remote collection to maintain control over the collection, while providing HEAL users with a central interface to search across many collections with one interface.

Utilizing a rich, consistent metadata standard in a multimedia digital library has a number of significant advantages.[16] A well designed metadata standard increases the effectiveness of end user searches and the likelihood that the end users will find what they are looking for. Users can search across multiple sites from one interface; in addition, the context of the item and its relationship to other items in the collection is retained when the metadata is downloaded along with the multimedia resource. Multiple versions of the same resource (for instance, a high resolution image and a low-resolution thumbnail) can be indexed with the same metadata. Well-structured metadata can also accomodate multiple user search strategies. Metadata also provides a mechanism for digital libraries to track usage rights and to document legal requirements such as privacy concerns (patient privacy rights, for instance) and intellectual property rights.

The HEAL Metadata Schema

Before developing the HEAL metadata schema, we researched a number of metadata standards developed by national and international standards organizations, including the Dublin Core, IEEE Learning Objects Metadata (LOM)[17], and Educause IMS[18]. The Dublin Core, originally designed to describe Web resources in a simple manner, consists of fifteen elements; we concluded that the Dublin Core was not robust enough to describe health sciences multimedia resources. Educause IMS, which consists of seventy nine elements and builds on both the Dublin Core and IEEE LOM, proved to be a better fit for the needs of the HEAL project.

Although the IMS elements capture most of the important information needed for describing a multimedia resource (e.g., author, title, keywords, copyright information, etc.), we determined that HEAL would need additional elements to reflect the unique needs of health science educators. We extended IMS to include additional health sciences related elements, including specimen type, radiograph type, orientation, magnification, annotations, disease process, and clinical history. Figure 5 illustrates the hierarchical relationship between Dublin Core, IEEE LOM, IMS, and the HEAL metadata standard. Because of the extended number of health sciences-specific elements the HEAL metadata standard allows users to precisely specify their search criteria.

In general, the breadth and depth of any metadata is subject to continual debate based on the variety of user perspectives and needs. The goal of the HEAL metadata specification is to find a balance between manageability and adequacy that both serves the broad needs of health sciences educators and the interoperability needs of the HEAL system. Although the HEAL metadata schema contains over seventy elements, the vast

majority of elements are entirely optional and that the required set of elements for any given item is quite minimal.

Other organizations may further extend the HEAL metadata standard if they need further specificity. To ensure that the HEAL collection is suitable for all levels of education the metadata schema includes information regarding educational level for which the resource is designed and flags materials that are inappropriate for minors. Although the current standard is designed to describe multimedia items, we anticipate that we will add elements to describe other learning object formats (e.g., lectures, cases, and quizzes) in the future.

When a faculty member contributes a resource to HEAL, he or she is required to provide basic metadata (e.g., author and copyright information, title of the resource, media type, suggested keywords and/or description of the resource, and educational rights). The suggested keywords are used to further catalog the resource using a controlled vocabulary; this is necessary in order to ensure that different keywords are not used to describe the same concept. HEAL resources are currently indexed with the Medical Subject Headings (MeSH) controlled vocabulary. MeSH consists of a hierarchical tree of health sciences related terms; this tree is available from the HEAL application for use by searchers and contributors.

While MeSH provides a good framework to begin indexing resources, it is often not specific enough to describe resources in many sub-specialties of medicine. To address this problem, we are in the process of investigating the possibilities offered by the Unified Medical Language System (UMLS)[19]. UMLS is a meta-thesaurus that maps

terminology from many different vocabularies to a central concept, including domain-specific vocabularies such as the Systematized Nomenclature of Medicine (SNOMED).

Faculty may currently contribute items to HEAL by completing a simple form that describes the contributed resource. In order to ensure that the item is indexed accurately and that the correct MeSH terms are employed, the HEAL team has hired a professional librarian to catalog resources. We believe that the use of professional catalogers is imperative to ensure the level of quality control and consistency we desire for the HEAL collection.

Broad acceptance from the health science education community of the HEAL metadata standard is crucial for facilitating the growth of the HEAL collection. We issued a Request For Comments (RFC) in February 2002 requesting feedback about our proposed metadata standard. The RFC was sent to the health science education community at large and to the organizations that had expressed interest in partnering with HEAL. The subsequent comments and revisions led to Version 1 of the HEAL metadata schema which was formally released in April 2002.

Components of the HEAL Metadata Schema

The HEAL core metadata elements are based on the content and structure of the widely-accepted Instructional Management Systems (IMS) metadata specification and were implemented using the eXetensible Markup Language (XML). They include i) selected elements from the IMS Metadata v.1.2 specification[20], and ii) several non-IMS elements that have utility within the Health Sciences (e.g. *SpecimenType*, *Organ*, etc.). The XML-based HEAL metadata schema follows the Best Practices recommended by the IMS Specification.[21]

The elements in the HEAL metadata schema (including the core elements based on IMS and the HEAL extensions) are organized into eight sections:

- Metametadata
- General
- Lifecycle
- Technical
- Educational
- Rights
- Relation
- Classification

These sections are described below. The HEAL metadata specification documentation and examples can be downloaded from the HEAL Web site.[22]

Metametadata. The metametadata section includes information about the catalog entry (a unique item and collection ID), the contributor of the resource, the language, and the version of the HEAL metadata schema. The contributor of the resource does not have to be the same as the Contributor who created the resource, which is described in the Lifecycle section.

General. The General section describes elements that can be used to describe the content of the resource. Elements adapted from IMS include Catalog Entry (unique identifiers for the collection and the item), Title and Description. This section also includes extensions specific for the health sciences, which are described in more detail below.

Lifecycle. Lifecycle contains contact information for the contributor, which may be described using the *vcard* standard for describing contact information, and the date of the contribution.

Technical. The technical sections describe the digital characters of the multimedia item, including the file extension, file height and width, running time (for videos and audios), file size, file location (URL), and software requirements. Since the resources in HEAL are Web resources by definition, the software requirements are used to describe particular browser versions and/or browser plug-ins required to view the item. The technical section also describes the format of the item. The format is a restricted vocabulary based on a partial list of the Ilumina formats.[23] Formats used in the HEAL metadata schema are: Audio, Animation, Chemical Structure, Executable, Image, Java Applet, Portable Document, Presentation, Video, Web Page, and Word Processing Document.

Educational. The educational section includes elements such as interactivity type, interactivity level, learning resource time, and typical learning time; these elements do not directly apply to the granular multimedia items available in HEAL, but are included for IMS compliance. The context element includes a description of the user group for which the multimedia item is designed. The restricted vocabulary for user group within the HEAL metadata schema is: K-12:primary; K-12:secondary; higher education, patient-education, and health profession education.

Rights. The rights section includes information about the copyright holder as well as a description of usage rights for the item. To be included in a HEAL, a rights statement must include, at a minimum, the right to freely use the materials for educational purposes. The copyright holder retains the copyright and may offer additional rights if desired. In

the future, the HEAL application will offer the contributor an option to use the Creative Commons tools[24] to develop a rights statement.

Classification. The classification section includes a taxon (controlled vocabulary or thesaurus) source and keywords. Examples of taxon source may include any standard controlled vocabulary such as MeSH or the Systemized Nomenclature of Medicine (SNOMED), or may reference a custom thesaurus developed by the contributor. More than one taxon source may be used in the HEAL metadata schema; the choice of a vocabulary source is determined by the particular needs of the subject areas described within the resource. For instance, contributors may recommend SNOMED rather than MeSH, or may specify a UMLS Concept Identifier to indicate the appropriate concept that can be mapped from many different vocabularies. Although keywords are not a required field, all items in HEAL are cataloged by a professional librarian using a controlled vocabulary.

HEAL Extensions. The IMS schema provided the basis for most of the HEAL metadata schema. While the IMS elements were adequate for the general and technical descriptions of the resource, we extended the schema to include elements specific to health sciences education, as well as elements that we felt were critical to meet the functional requirements of the HEAL system.

The health science related extensions were based on our experience with a previous health sciences multimedia database used to describe videodisc images[25]. The health sciences specific elements in the HEAL metadata schema are described in Table 1.

Table 1. Health Sciences Extensions

Element Name	Format	Description	Values
Specimen Type	Standard Pick List	Restricted vocabulary: {cell, tissue, organ, organ system; organelle}	“organ system”
Radiograph Type	Standard Pick List	Radiology technology used to generate the multimedia item. Restricted vocabulary: angiogram, CT, MRI, nuclear, PET, Plain, ultrasound	"MRI"
Magnification	Free text	Magnification of microscopic image	“100x”
Disease Process	Standard Pick List; Multiplicity allowed	Indicates disease process displayed, discussed, or implied in the item. Restricted vocabulary: vascular, inflammatory, neoplastic, degenerative, intoxication, congenital, allergic/autoimmune, traumatic, endocrine, social, legal, environmental, economical, psychological	"endocrine"
Clinical History	Free text; Multiplicity allowed	Clinical history of patient	“Fifty-year old male with family history of congestive heart failure presented with acute chest pain and shortness of breath.”

In addition to the health sciences specific extensions, we identified elements that were needed to satisfy the functional requirements of the HEAL system. These elements are described in Table 2.

Table 2. HEAL Extensions to Satisfy Functional Requirements

Element Name	Format	Description	Values
Annotated	Boolean ; no is the default value	“Yes” indicates that the multimedia is labeled or annotated; “No” indicates no annotations	"yes"
Inappropriate for minors	Boolean : default value is “No”	Used to flag materials inappropriate for minors	"no"
Context URL	Free text	URL pointing to the educational context in which the item is used . The context in which the item is used can be a course, a case, etc. Contributor can add this information if he/she believes that the context is essential for the correct understanding/usage of the item.	“http://www.medsch.ucla.edu/histology”
Context URL Description	Free text	Description of the ContextURL	This web site provides additional narrative text that explains the structures on this image.

Extending the HEAL Metadata Schema

The XML-based HEAL metadata schema was designed to describe a variety of common multimedia items (e.g. images, animations, videos, etc.). This core metadata may also, to a limited extent, be used to describe more complex resources such as problem-based learning cases, practice quiz questions, etc. Future versions of HEAL

metadata will provide additional metadata elements aimed at these more complex resources.

The element definitions are not necessarily bound to a particular implementation technology; organizations are free to implement the HEAL metadata specification through a variety of means including database and XML applications. Technical implementers should note that the HEAL-specific (i.e. non-IMS) elements have been added according to the IMS mechanism for extension outlined in the IMS Learning Resource Meta-data Best Practice and Implementation Guide. Locally-developed applications may further extend this HEAL specification through this same mechanism.

Future Work

According to the HEAL Web site logs, the HEAL metadata schema has been downloaded by over 300 users since its release. Over a half a dozen institutions that are implementing local versions of a multimedia resource database have adapted the schema for local use. We are currently collecting feedback from our users and plan to add a metadata registry on our Web site so that we may track downloads and follow up with users of the metadata schema. We anticipate releasing an updated version based on this feedback by the fall of 2004. In addition, we plan to extend the HEAL metadata schema so that is suitable for indexing other types of resources, including interactive cases and quizzes.

Conclusion

HEAL has been designed to facilitate the sharing of and improve the access to a wide variety of high-quality, freely available multimedia resources that are currently

located across many remote servers. This application provides educators with a single but powerful search interface through which they can simultaneously search multiple collections for high quality materials that will enhance their teaching.

The HEAL metadata schema is the foundation on which the HEAL distributed system is built; the schema is based on international standards and includes extensions specific to the health sciences. Users may implement the HEAL metadata schema on their local systems or may further extend the schema to meet local needs. The metadata schema provides an important mechanism for the HEAL central server to share data with partner collections located on remote servers. As feedback is received from the health sciences community, the HEAL metadata schema will be improved and updated. HEAL will have a positive impact on the health education community as the team is motivated by an open, collaborative spirit and is determined to share its materials widely; as such, all content, metadata, standards, and application code are made freely available to all audiences and users.

This project has been made possible by grants from the National Science Foundation, DUE-0085660 and DUE-0226314, and from the National Library of Medicine, 1 G08 LM007877-01.

Figures

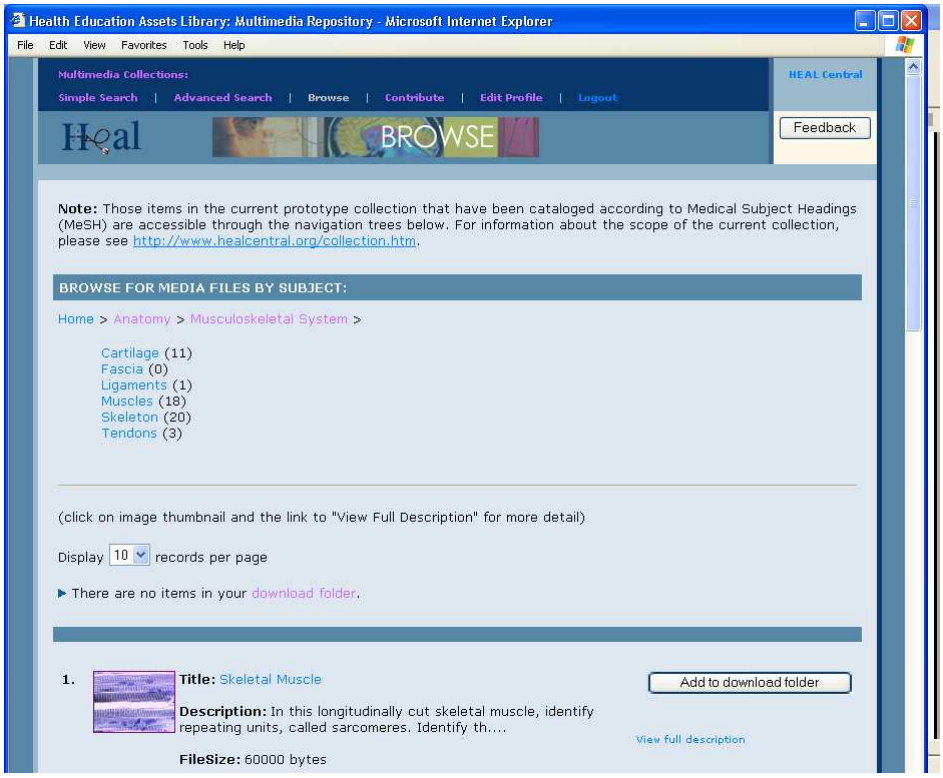

Figure 1. The Browse interface allows users to peruse the collection via the hierarchical MeSH tree. The number of items contained within each heading is indicated in parentheses.

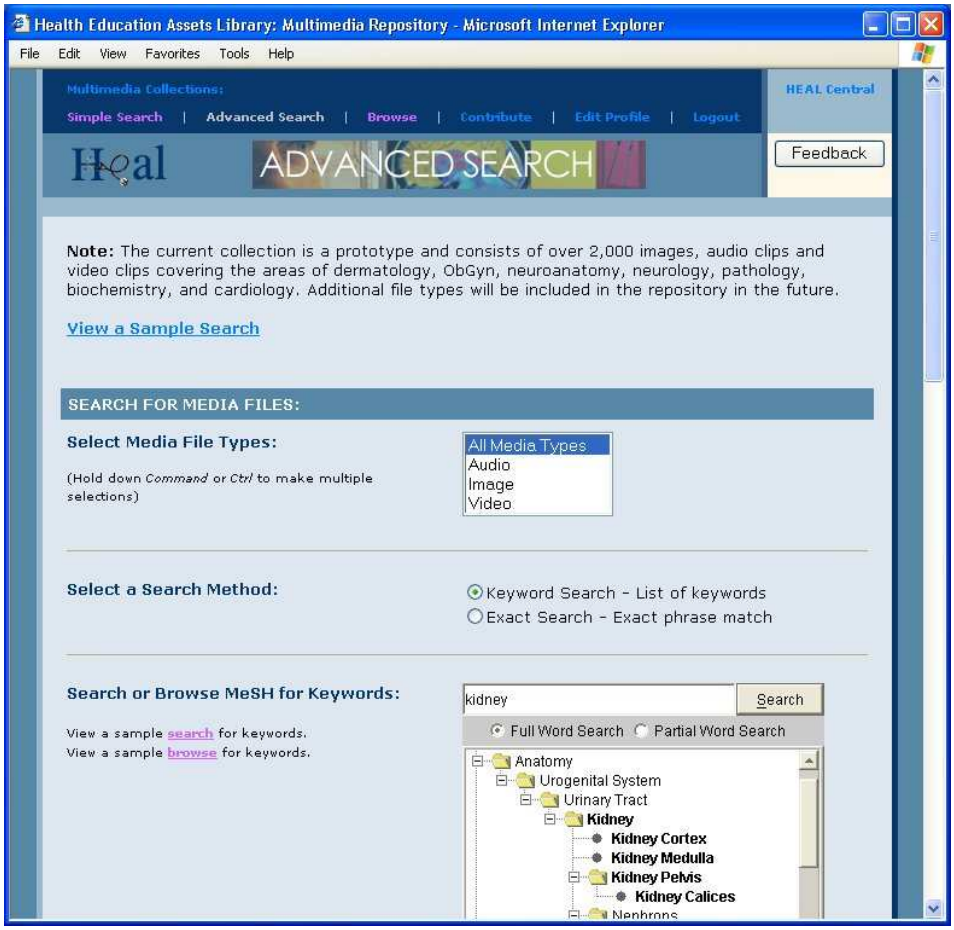

Figure 2. The Advanced Search interface allows users to conduct precise searches using a combination of specific search criteria.

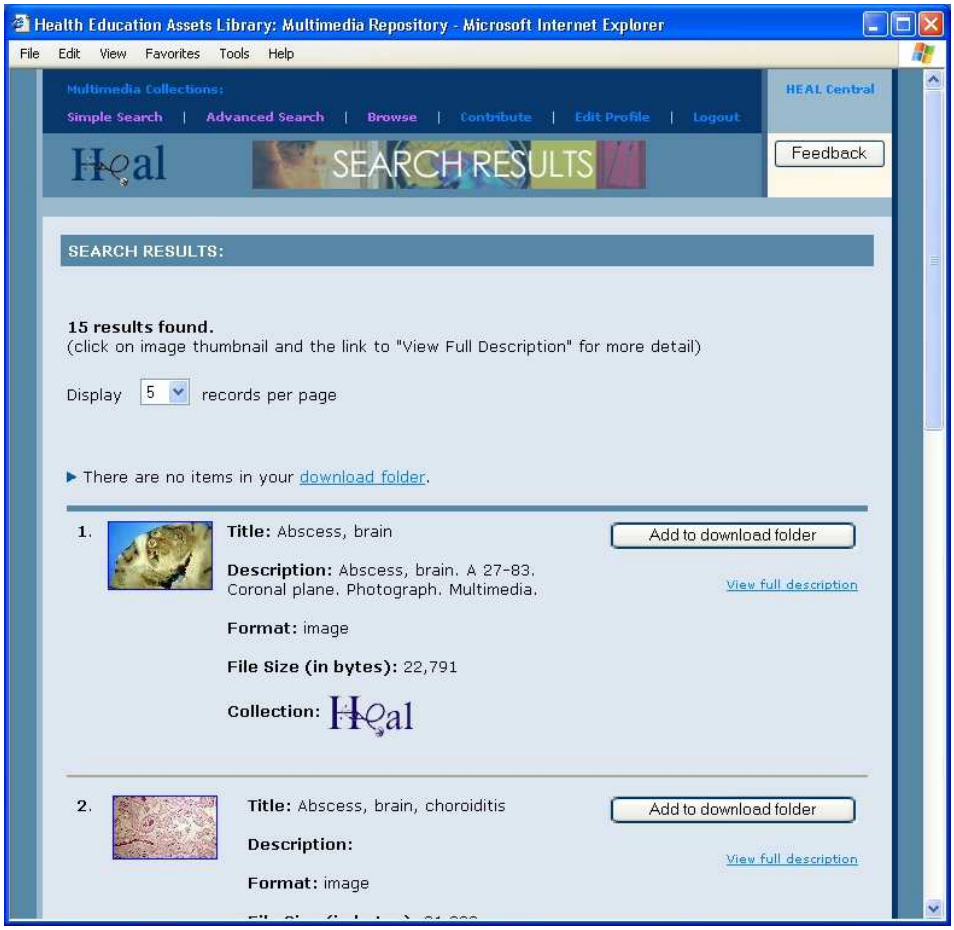

Figure 3. The Search Results interface presents a thumbnail sketch of each item in addition to a description, the source collection, the format, and the file size**.**

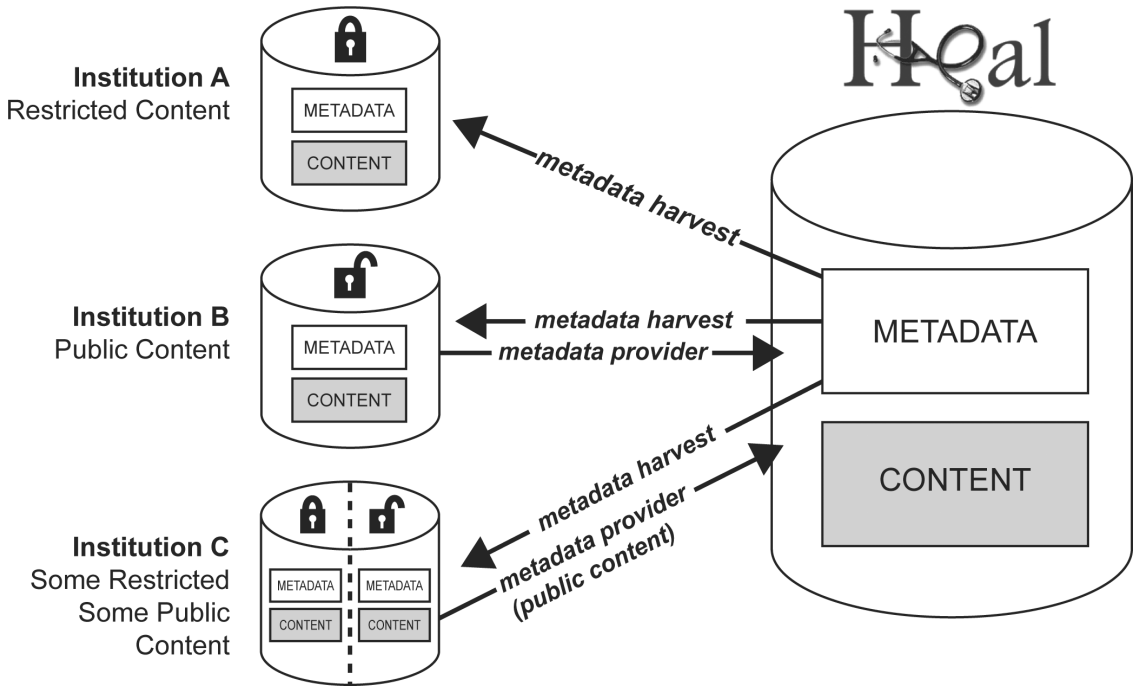

Figure 4: The HEAL metadata harvesting system. Using the Open Archives Initiative (OAI), institutional collections can elect to share all, none, or a portion of their catalog metadata with the central HEAL server, all while obtaining the HEAL catalog metadata for their own uses.

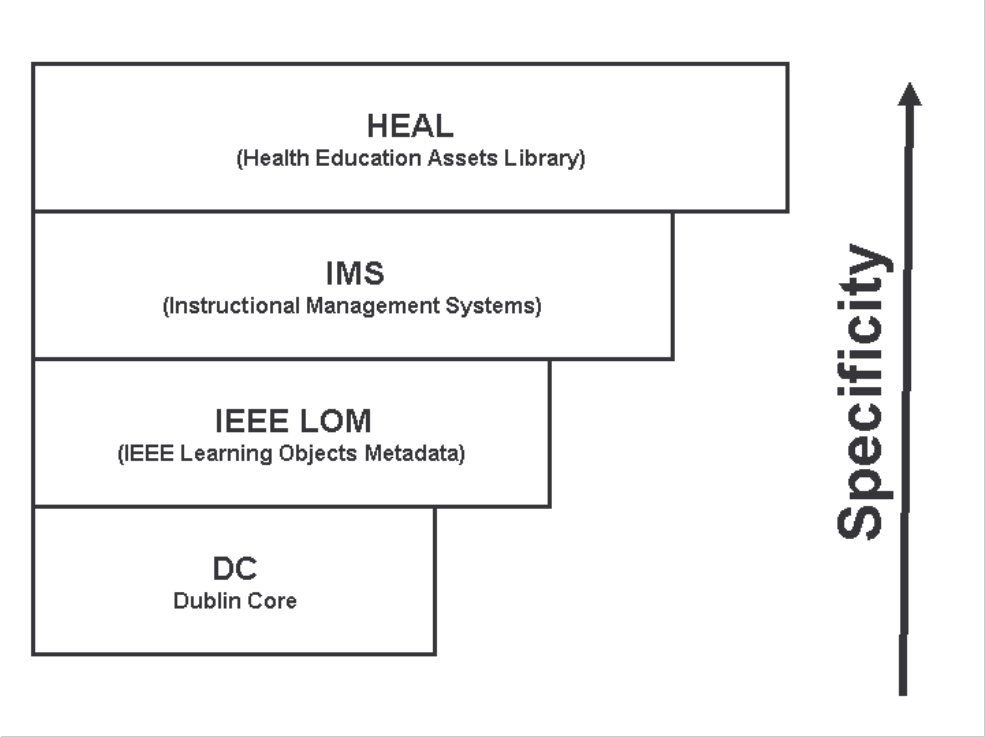

Figure 5: The hierarchical relationship of the HEAL metadata schema compared to other common metadata schemas. Each layer extends the subset of its child schema through additional elements which provide more specificity.

[1] Klatt, E.C. and Dennis, S.E. Web-based Pathology Education. *Arch Pathol Lab Med*; 122(5):475-479, May 1998.

[2] Wiemer W. It's a Long Way to Multimedia: An Account of 18 Years of Pursuing a New Media Project in Physiology. 1998. Advances in Physiology Education. 1998;20:S96-S105

[3] Frisse ME. The Case for Hypermedia. 1990; 65: 17 – 19.

[4] Henry J. Computers in Medical Education: Information and Knowledge Management, Understanding, and Learning. Human Pathology. 1990;21:998-1002.

[5] Moberg TF, Whitcomb ME. Educational Technology to Facilitate Medical Student's Learning: Background Paper 2 of the Medical School Objectives Project. *Academic Medicine*. 1999;74:1145-1150.

[6] Greenhalgh T. Computer assisted learning in undergraduate medical education. *BMJ*. 2001;322:40-44

[7] Todorovski L, Ribaric S, Dimec J, Hudornalj E, Lunder T. Organization and dissemination of multimedia medical databases on the WWW. Medical Informatics Europe. IOS Press. 1999. 557 - 561

[8] Candler CS, Andrews MD. Avoiding the great train wreck: standardizing the architecture for online curricula. *Academic Medicine*. 1999; 74: 1091-1095

[9] National Science Digital Library: http://www.nsdl.org [Accessed June 30, 2003]

[10] Candler C, Uijtdehaage S, Dennis S. Introducing HEAL: the Health Education Assets Library. A*cademic Medicine*. 2003; March 1, 2003; 78(3): 249-253

[11] Health Education Assets Library: http://www.healcentral.org [Accessed June 30, 2003].

12 Open Archives Initiative: http://www.openarchives.org [Accessed June 30, 2003].

[13] Introduction to Vocabularies, Getty Research Institute, http://www.getty.edu/research/institute/vocabulary/introvocabs/ [Accessed June 30, 2003]

[14] Weibel, S, Lagoze, C. An element set to support resource discovery: the state of the Dublin Core. *International Journal on Digital Libraries*. January 1997, 1(2): 176-186.

[15] Dublin Core Metadata Initiative, http://www.dublincore.org/ [Accessed June 30, 2003]

[16] Introduction to Metadata, Getty Research Institute, http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html, [Accessed June 30, 2003]

[17] IEEE Learning Technology Standards Committee, http://ltsc.ieee.org/wg12/ [Accessed June 30, 2003]

18 IMS Global Learning Consortium: http://www.imsproject.org [Accessed June 30, 2003].

[19] Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc 1997 Nov-Dec;4(6):484-500

[20] IMS Learning Resource Meta-data Model, Version 1.2 Final Specification http://www.imsproject.org/metadata/imsmdv1p2/imsmd_infov1p2.html [Accessed June 30, 2003]
[21] IMS Learning Resource Meta-data Best Practice and Implementation Guide: http://www.imsproject.org/metadata/imsmdv1p2/imsmd_bestv1p2.html [Accessed June 30, 2003]
[22] HEAL Documents: http://www.healcentral.org/documents.htm [Accessed June 30, 2003]
[23] Ilumina: http://turing.bear.uncw.edu/iLumina/index.asp [Accessed June 30, 2003]
[24] Creative Commons: http://www.creativecommons.org [Accessed June 30, 2003]
[25] Stensaas, S. Animating the curriculum: integrating multimedia into teaching. Bull Med Libr Assoc. 1994 Apr;82(2):133-9.